Contribution to GISC project: Flagging data quality

Leonor Tarrasón

AQ and Atmospheric Composition Metadata Workshop, Dublin, Ireland
5th-7th September 2012

GISC project support by ETC/ACM partners

DELIVERABLE REPORT
TASK 3

Harmonised
quality flagging
of air quality data

Solutions for meeting the recommendations
from GMES atmospheric core service by EIONET networks

FIRST DRAFT
24.02.2012

NILU

INERIS

European Environment Agency
GISC
GMES In-Situ Coordination

- *Recommendations for harmonisation QA/QC data flagging of in-situ data*

- *From the perspective of model users (GMES/MACC service providers of AQ forecasts)*

*Authors: Leonor Tarrasón, Aasmund Fahre Vik, NILU,
Laurence Rouïl, Laure Malherbe and Anthony Ung, INERIS*

Structure of the report

- **Main principles of GMES AQ flagging System**

- Implications for EIONET

- Implications for national data providers (Norway, France)

- Implications for research networks

- Recommendations

Modelling perspective

- Use of observation data as basis for data assimilation and/or for validation purposes
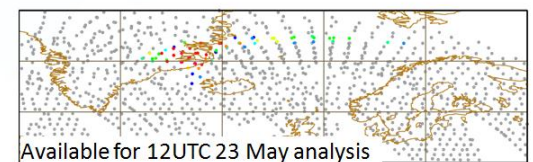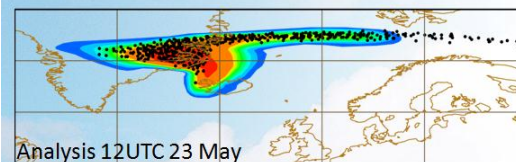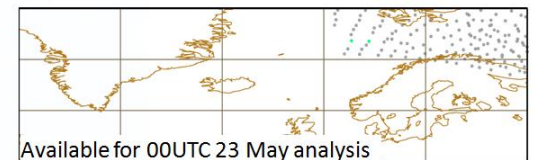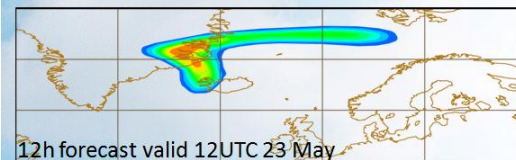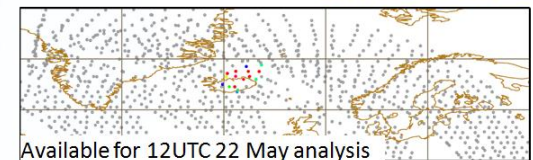
3 types of data

- Near real time data
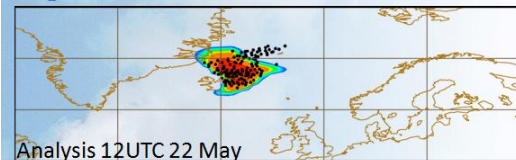- Delayed mode data
- Validated data

GRG/AER: Volcanic eruptions (Grimsvötn, May 2011)

SO_2 Model values and IASI detection — OMI observations

Analysis 12UTC 22 May — Available for 12UTC 22 May analysis

12h forecast valid 12UTC 23 May — Available for 00UTC 23 May analysis

Analysis 12UTC 23 May — Available for 12UTC 23 May analysis

2 3 4 5 6 7 8 9 10 20 50 100
Dobson units

Validation Stage as part of data description
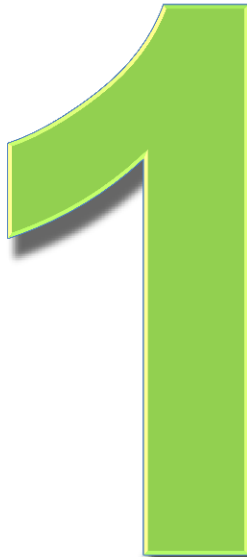
The data flagging system needs to be appllicable to

    -Near real time data
   -Data supplied in delayed mode
    -Validated data

The aim is that NRT and validated data are stored and accessed in a consistent way

Station uncertainty and characterisation

Relevant infomation for use of observation data in model applications

➢Evaluation of station representativeness
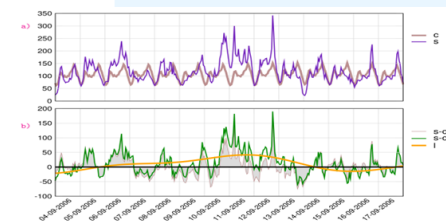➢Station uncertainty -Feedback from modelling users to data providers

Case study for Polish station: Dabrowa Gornicza

20 km buffer

10 km buffer

5 km buffer
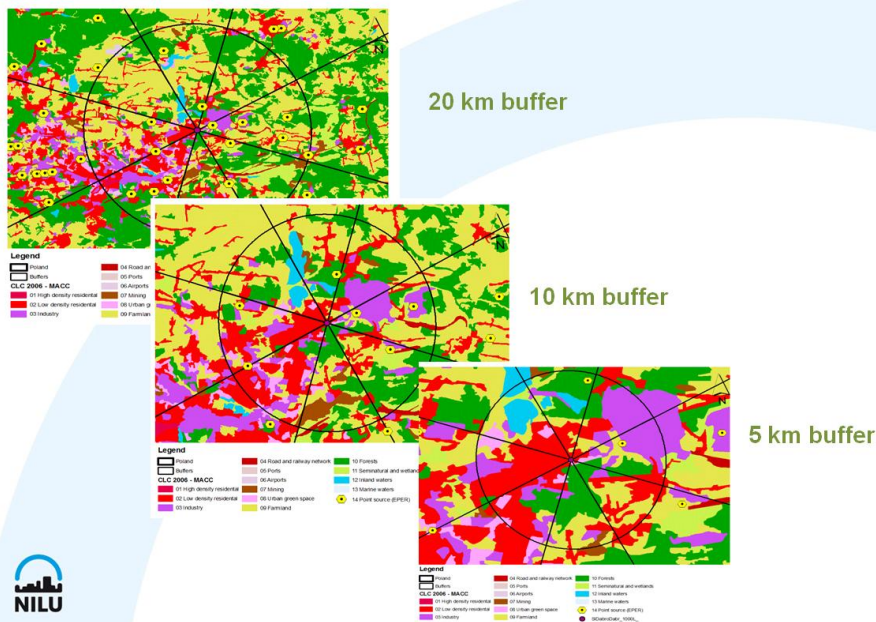
Classification in 10 groups vs Airbase site metadata
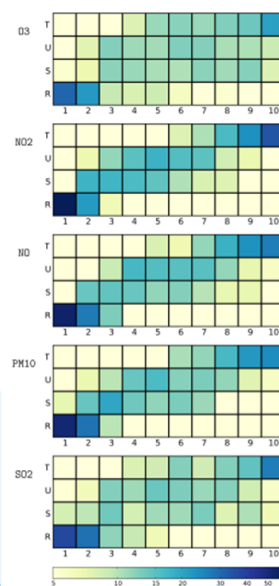
The station « real fingerprint » is to be found in its past observations rather than in external information - Peuch and Joly, 2010 -

	AIRBASE Site	AIRBASE Area	BDQA	O3	NO2	NO	PM10	SO2
T	Traffic	Urban	Traffic	187	475	382	328	243
U	Background	Urban	Urban	496	616	433	462	400
S	Background	Suburban	Suburban	268	256	224	182	137
R	Background	Rural	Rural	293	153	127	106	128
O	Others	Others	Others	374	561	459	416	491
Total				1618	2061	1625	1494	1399

3 different types of metadata

Type I : Static metadata: external properties, provenance and ownership – like station position *Can be stored separatedly from the actual data.*

Type II : Stable metadata : Data Description – This is the part of the data description that may change regularly: like information on teh station representativeness feedback from modellers on the uncertainty of the observations

Type III: Varaible metadata: *Data description that needs to be transmitted and stored with the actual data.* This includes especially information that characterizes the instrument and method used to carry out the actual observation.

Data description: characterisation of data

- **Quality control flag (QC) - Validation Stage Flag**

Flag	Quality control level	Comments for external use
1	Raw data	Data directly from the data log
2	Automatically controlled data	Data check is done automatically when the data gets into the reporting database.
3	Calibrated data	The data is calibrated as a function of instrument used. – Special automatic module "Scaling"
4	Validated data	The data is manually checked and validated data is flagged with the use of the module "Approval"

Already been recognised by the European Commission under the work with new implementing provisions (IPR) of the Air Quality Directive 2008/50/EC (AQD).

Data description: characterisation of data

- ## Quality Assurance (QA) flag

The QA flag determines whether a measurement value is valid or not.

Each individual measurement has a separate quality assurance flag.

QA flags are generated by combining
- ▪ the instrument flag (activated if there are failures from instrument) and
- ▪ the automatic control system including data capture.

Instrument: TEOM

MODE Flag (last digit, first from the right):

Flag	Description
1	The instrument has been initialized
2	Data collection has begun
3	Total mass computed, but no mass concentration.
4	Normal operation

STATUS Flag (position 2 and 3 from the right):

Flag	Description
Blank	OK (normal)
1	Filter error
2	Flow error
3	Flow and filter error
4	Temperature error
5	Temp and filter error
6	Temp and flow error
7	Temp and flow and filter error
8	Frequency error
9	Frequency and filter error
10	Frequency and flow error
11	Frequency and flow and filter error
12	Frequency and temp error
13	Frequency and temp and filter error
14	Frequency and temp and flow error
15	Frequency and temp and flow and filter error

Note: Error Flags10-15 are denoted as A-F in the TEOM datalog, but NILU converts these to numbers in order to achieve a more homogeneous visualization of instrument error in the website.

Revision of QA/QC flagging

The new IPR proposes an extension of the flagging system to:

- valid
- valid, but value is below detection limit, number not replaced
- valid, but value is below detection limit , number replaced by 0.5*detection limit
- not valid due to station maintenance or calibration
- not valid due to other reasons or missing

The proposal from this work is to include an additional flag:

- **valid, but value replaced after calibration**

This flagging system is part of the IPR guidance document which is currently being discussed by Member States for the revision of the AQD

Thank you for your attention!